

## Ethics, Diversity and Consciousness in AI



**E**thics in Artificial Intelligence (AI) is discussed everywhere, in governmental circles as well as scientific forums. It is very often associated with the two widely used concepts of diversity and eXplainable Artificial Intelligence (XAI). The latter was promoted by DARPA<sup>1</sup>, and it is expected to propose methods preserving the rights of users to understand how the AI systems work and why decisions are made. Computational Intelligence methods have much to contribute to this effort to ensure that AI systems are ethical.

In July 2020, the European Union has published an Assessment List for Trustworthy Artificial Intelligence (ALTAI)<sup>2</sup> covering a wide range of ethical issues. Among the requirements listed, we find transparency, that encompasses the traceability of the data and model used to make a decision, the explainability of decisions and the communication with the user. Transparency mainly corresponds to the general approach of XAI and Computational Intelligence has much to do to participate in the effort of transparency of XAI, by means of the various methods of automatic generation of explanations, the use of counterfactuals, the help of visualization or the construction of fuzzy systems providing interpretability and expressiveness, for example.

Another ALTAI requirement sounds familiar to us, dealing with diversity, non-discrimination and avoidance of unfair bias. Such concerns are at the heart of the IEEE Computational Intelligence's actions, which is very committed to their respect in its scientific and associative life. It is therefore natural to also consider these properties as essential for all algorithms and products that are parts of AI systems, not only in the collection of data and choice of attributes leading to a decision, but also in the ability of all categories of users to interact with the AI system regardless of their characteristics. Privacy and data protection is another aspect of the respect of human users.

Other requirements focus more on the cognitive components of the construction of AI systems, the effect they can have on human users, the possible governance of human agents in AI systems and the integration of AI systems in the human society. The technical robustness of the AI system, its reliability, its security and its capacity to identify and mitigate risks in a transparent way are not less important.

Machine ethics and the notions of responsibility of AI systems or artificial moral agents<sup>3</sup> are also topics of research that deserve a lot of attention. Consciousness in AI has been debated for a long time<sup>4</sup> and it can take a form different from human consciousness<sup>5</sup>. Multidisciplinary, and in particular the cooperation with cognitive

### CIS Society Officers

*President – Bernadette Bouchon-Meunier, Sorbonne Université, FRANCE*  
*President Elect – Jim Keller, University of Missouri, USA*  
*Vice President – Conferences – Marley M. B. R. Vellasco, Pontifical Catholic University of Rio de Janeiro, BRAZIL*  
*Vice President – Education – Pau-Choo (Julia) Chung, National Cheng Kung University, TAIWAN*  
*Vice President – Finances – Pablo A. Estévez, University of Chile, CHILE*  
*Vice President – Members Activities – Sanaz Mostaghim, Otto von Guericke University of Magdeburg, GERMANY*  
*Vice President – Publications – Kay Chen Tan, Hong Kong Polytechnic University, HONG KONG*  
*Vice President – Technical Activities – Luis Magdalena, Universidad Politécnica de Madrid, SPAIN*

### Publication Editors

*IEEE Transactions on Neural Networks and Learning Systems*  
*Haibo He, University of Rhode Island, USA*  
*IEEE Transactions on Fuzzy Systems*  
*Jon Garibaldi, University of Nottingham, UK*  
*IEEE Transactions on Evolutionary Computation*  
*Carlos A. Coello Coello, CINVESTAV-IPN, MEXICO*  
*IEEE Transactions on Games*  
*Julian Togelius, New York University, USA*  
*IEEE Transactions on Cognitive and Developmental Systems*  
*Yaochu Jin, University of Surrey, UK*  
*IEEE Transactions on Emerging Topics in Computational Intelligence*  
*Yew Soon Ong, Nanyang Technological University, SINGAPORE*  
*IEEE Transactions on Artificial Intelligence*  
*Hussein Abbass, University of New South Wales, AUSTRALIA*

### Administrative Committee

#### Term ending in 2021:

*David Fogel, Natural Selection, Inc., USA*  
*Barbara Hammer, Bielefeld University, GERMANY*  
*Yonghong (Catherine) Huang, McAfee LLC, USA*  
*Xin Yao, Southern University of Science and Technology, CHINA*  
*Jacek M. Zurada, University of Louisville, USA*

#### Term ending in 2022:

*Cesare Alippi, Politecnico di Milano, ITALY*  
*James C. Bezdek, USA*  
*Gary Fogel, Natural Selection, Inc., USA*  
*Yaochu Jin, University of Surrey, UK*  
*Alice E. Smith, Auburn University, USA*

#### Term ending in 2023:

*Piero P. Bonissone, Piero P Bonissone Analytics LLC, USA*  
*Oscar Cordon, University of Granada, SPAIN*  
*Pauline Haddow, Norwegian University of Science and Technology, NORWAY*  
*Haibo He, University of Rhode Island, USA*  
*Hisao Ishibuchi, Southern University of Science and Technology, CHINA*

<sup>1</sup>[https://www.cc.gatech.edu/~alanwags/DLAI2016/\(Gunning\)%20IJCAI-16%20DLAI%20WS.pdf](https://www.cc.gatech.edu/~alanwags/DLAI2016/(Gunning)%20IJCAI-16%20DLAI%20WS.pdf)

<sup>2</sup><https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai>

<sup>3</sup><https://plato.stanford.edu/entries/ethics-ai/#MachEthi>

<sup>4</sup>McCarthy, John (1996). Making robots conscious of their mental states. In S. Muggleton (ed.), *Machine Intelligence 15*. Oxford University Press.

<sup>5</sup>Pitrat, Jacques (2009). *Artificial Beings : The Conscience of a Conscious Machine*. Wiley, 2009

## A special issue of the IEEE Computational Intelligence Magazine dedicated to Explainable and Trustworthy Artificial Intelligence will be published in 2022.

sciences and computational neuroscience, is important in this regard.

The Computational Intelligence society (CIS) is at the heart of all these aspects of ethics in AI. They represent fascinating challenges for researchers and they can highlight the power of all Computational Intelligence methods: neural networks, learning methods, fuzzy systems and evolutionary computation. Efforts are already done through the IEEE CIS Task Force on Ethical and Social Implications of Computational Intelligence. Several other CIS

Technical Committee task forces and special issues of CIS publications are also focused on XAI. In particular, a special issue of the IEEE Computational Intelligence Magazine dedicated to Explainable and Trustworthy Artificial Intelligence will be published in 2022. The IEEE CIS Cognitive and Developmental Systems Technical Committee and the IEEE Transactions on Cognitive and Developmental Systems also address some of these concerns on a cognitive basis. Moreover, the IEEE CIS is the oversight committee for the

IEEE Consortium On The Landscape of AI Safety (CLAIS) and, on a more specific level, it participates in the IEEE Brain Community.

I am convinced that the IEEE CIS needs to pay even more attention to these crucial issues. At a time when the IEEE CIS is strengthening its ties to industry by creating a Committee for Industrial and Governmental Activities starting in January 2022, the trustworthiness of Artificial Intelligence appears even more as a key factor in the realization and success of industrial products based on Computational Intelligence methods.

If you are interested in these topics or any other activity of the IEEE CIS, you can always contact me at [b.bouchon-meunier@ieee.org](mailto:b.bouchon-meunier@ieee.org).



# TAP. CONNECT. NETWORK. SHARE.



## Connect to IEEE—no matter where you are—with the IEEE App.



Stay up-to-date  
with the latest news



Schedule, manage, or  
join meetups virtually



Get geo and interest-based  
recommendations



Read and download  
your IEEE magazines



Create a personalized  
experience



Locate IEEE members by location,  
interests, and affiliations

**Download Today!**

