Principles for the Trustworthy Adoption of AI in Legal Systems:

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems

Nicolas Economou
H5
340 Madison Avenue
19th Floor
New York, NY USA 10173
neconomou@h5.com

ABSTRACT

The advent of artificial intelligence in legal systems spurred laudable efforts to assess its implications, risks, and benefits. Among those efforts, US NIST's TREC Legal Track produced exemplary scholarship on the effectiveness of AI in discovery; other initiatives explored bias in risk-assessment algorithms used in bail or sentencing; and bar associations considered the implications for professional conduct. Yet, a foundational question remained unaddressed: What framework could equip lawyers, judges, advocates, policy makers, and the public, irrespective of legal system or cultural traditions, to determine the extent to which they should trust (or mistrust) the deployment of AI in the legal system? The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, a multiyear, international, multidisciplinary effort focused on the ethics of AI took on this challenge. This talk, by the Chair of the Initiative's Law Committee, will present the IEEE's recently published proposed norms for the trustworthy adoption of AI in legal systems, outline the objectives of its upcoming work, and place this endeavor in the broader context of international law-focused AI governance endeavors.

In: Proceedings of the First International Workshop on AI and Intelligent Assistance for Legal Professionals in the Digital Workplace (LegalAIIA 2019), held in conjunction with ICAIL 2019. June 17, 2019. Montréal, QC, Canada.

Keynote Presentation

The advent of artificial intelligence in legal systems since the early 2000s spurred laudable efforts to assess its implications, risks, and benefits. Among those, US NIST's seminal TREC Legal Track studies produced exemplary scholarship on the effectiveness of AI in discovery. Several initiatives explored bias in risk assessment algorithms used in bail or sentencing. Bar associations considered the implications for professional conduct. Yet, a foundational question remained unaddressed: what framework and instruments could equip lawyers, judges, advocates, policy makers, and the public, irrespective of legal system or cultural traditions, to determine the extent to which they should trust (or mistrust) the deployment of AI in legal systems.

The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems, a multiyear, international, multidisciplinary effort focused on the ethics of Artificial Intelligence took on this challenge. The IEEE, which traces its roots back to Thomas Edison and Alexander Graham Bell, is a global technology think tank and one of the world's leading standards-setting bodies. The IEEE Global Initiative's mission is "to ensure every stakeholder involved in the design and development of autonomous and intelligent systems is educated, trained, and empowered to prioritize ethical considerations so that these technologies are advanced for the

benefit of humanity." In early 2019, the Global Initiative published its treatise, *Ethically Aligned Design*, First Edition ("EAD") which sets forth the high-level ethical principles, key issues, and recommendations to advance this mission.

When it comes specifically to the trustworthy adoption of Artificial Intelligence in legal systems and the practice of law, the IEEE Global Initiative's Law Committee sought to answer this central question: "When it comes to legal systems, to what extent should society delegate to intelligent machines decisions that affect people?"

The IEEE Law Committee EAD Chapter proposes that a definition of "**Informed Trust**" is necessary in order to answer this question and that this definition must meet certain design constraints. Specifically, it needs to rest on a single set of principles that are:

- Individually necessary and collectively sufficient
- Applicable to the totality of the legal system
- Globally applicable but culturally flexible
- Considering the legal system as an institution accountable to the citizen (so as to avoid solely considering professional ethics or judicial ethics, etc.)
- Capable of being operationalized

The IEEE Law Committee concluded that four principles fulfill the above design conditions in defining "Informed Trust" in the adoption (or avoidance of adoption) of AI in legal systems and the practice of law:

1. Effectiveness
2. Competence
3. Accountability
4. Transparency

Those principles are outlines below.

Principle 1: Evidence of Effectiveness

An essential component of trust in a technology is trust that it in fact works and succeeds in meeting the purpose for which it is intended. The principle of effectiveness, by requiring the collection and disclosure of evidence of the effectiveness of AI-enabled systems applied to legal tasks, is intended to ensure that stakeholders have the information needed to have a well-grounded trust that the systems being applied can meet their intended purposes. In order for the practice of measuring effectiveness to realize its potential for fostering trust and mitigating the risks of uninformed adoption and uninformed avoidance of adoption, it must have the certain features: Meaningful metrics that are practically feasible and actually implemented; Sound methods. Valid data; Awareness and consensus; Transparency.

Principle 2: Competence

An essential component of informed trust in a technological system, especially one that may affect us in profound ways, is confidence in the competence of the operator(s) of the technology. We trust surgeons or pilots with our lives because we have confidence that they have the knowledge, skills, and experience to apply the tools and methods needed to carry out their tasks effectively. We have that confidence because we know that these operators have met rigorous professional and scientific accreditation standards before being allowed to step into the operating room or cockpit. This informed trust in operator competence is what gives us confidence that surgery or air travel (or even a plumbing repair!) will result in the desired outcome. No such standards of operator competence currently exist with respect to AI applied in legal systems, where the life, liberty, and rights of citizens can be at stake. Such standards are both indispensable and considerably overdue.

Principle 3: Accountability

An essential component of informed trust in a technological system is confidence that it is possible, if the need arises, to apportion responsibility among the human agents engaged along the path of its creation and application: from design through to development, procurement, deployment, operation, and, finally, validation of effectiveness. Unless there are mechanisms to hold the agents engaged in these steps accountable, it will be difficult or impossible to assess responsibility for the outcome of the system under any framework, whether a formal legal framework or a less formal normative framework. A model of AI creation and use that does not have such mechanisms will also lack important forms of deterrence against poorly thought-out design, casual adoption, and inappropriate use of AI.

Principle 4: Transparency

An essential component of informed trust in a technological system is confidence that the information required for a human to understand why the system behaves a certain way in a specific circumstance (or would behave in a hypothetical circumstance) will be accessible. Without appropriate transparency, there is no basis for trusting that a given decision or outcome of the system can be explained, replicated, or, if necessary, corrected. Without appropriate transparency, there is no basis for informed trust that the system can be operated in a way that achieves its ends reliably and consistently or that the system will not be used in a way that impinges on human rights. In the case of AI applied in a legal system, such a lack of trust could undermine the credibility of the legal system itself.

An effective implementation of the transparency principle will ensure that the appropriate information is disclosed to the appropriate stakeholders to meet appropriate information needs, striking a balance between legitimate grounds for withholding information (privacy, security, intellectual property) and the needs of a legitimate inquiry into the design and operation of an AI-enabled system.

Next steps – From Principles to Practice

With these principles established, the IEEE will seek to develop instruments, such as standards and certifications, which can serve as the "Currency of Trust", which lawyers, judges, procurement officers, policy makers, advocates and the public can understand in determining the extent to which AI-enabled systems and their operators meet certain criteria or claims. In this regard, the IEEE has established The Ethics Certification Program for Autonomous and Intelligent Systems, which will progressively develop such instruments.

It should be noted that, independently but nearly simultaneously to the IEEE's work, the Council of Europe published the first Ethical Charter promulgated by an intergovernmental organization for use of Artificial Intelligence in judicial systems and their environment. The prominence of the Council of Europe renders this work of particular importance to stakeholders in legal systems globally. The Council of Europe, in the context of an international multi-stakeholder roundtable on AI and the Rule of Law recently launched a project for the certification of artificial intelligence in the light of the Charter, further strengthening the global impetus for trustworthy norms for AI in the law.

About the Author

Nicolas Economou is the chief executive of H5 and was a pioneer in advocating the application of scientific methods to electronic discovery. He chairs the Law Committees of the IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems and of the Global Governance of AI Roundtable hosted in Dubai as part of the annual World Government Summit. He leads The Future Society's Law Initiative and is a member of the

Council on Extended Intelligence (CXI), a joint initiative of the MIT Media Lab and IEEE-SA. He has spoken on issues pertaining to artificial intelligence and its governance at a wide variety of conferences and organizations, including the Spring Meetings of the International Monetary Fund (IMF), UNESCO, Harvard and Stanford Law Schools, and Renmin University of China. Trained in political science at the Graduate Institute of International Studies of the University of Geneva (Switzerland), he earned his M.B.A. from the Wharton School of Business, and chose to forgo completion of his M.P.A at Harvard's Kennedy School in order to co-found H5.