Hans Franken Lecture   'Artificial Intelligence and Human Rights'
by Jan Kleijssen, 30 June 2023

Two weeks ago, it was reported that the congregation in the fully packed Protestant Church of St Paul's in the Bavarian town of Fuerth were asked to raise and praise the Lord.

Then they were told by an avatar of a bearded man in black on a huge screen: "Dear friends, it is an honour for me to stand here and preach to you as the first artificial intelligence at this year's convention of Protestants in Germany."

The entire Service was created, as you will have guessed, by Chat GTP, with the help of an Austrian theologian.

Reactions were, as will not surprise you, very mixed.

As Yuval Noah Harari observed in his recent essay in the Economist; "Religions throughout history have claimed a non-human source for their holy books. Soon that might be a reality".

And not only holy books, but very concretely also the promises they hold. Such as immortality.

In China, cemeteries, and private individuals are resuscitating the deceased by creating AI made avatars, who look, speak and indeed, seem to think, like the cherished departed.

I invite you to reflect for a moment on the ethical and legal implications, as well as the risks, of such an approach.

If we were to put our faith in Chat GTP , it would appear that this practice has also reached the Netherlands.

Because in response to my query about our illustrious host, Professor Hans Franken, Chat GTP not only informed me of his many outstanding achievements, but also stated that he had sadly passed away in 2011.

It then added, on an optimistic note, that there might have been career developments since, of which it might not be aware.

So either we are indeed fortunate to be in the presence of a superb avatar today and to have witnessed its remarkable achievements for the past 12 years or was I confronted with one of Chat GTP's famous "hallucinations"?

Actually, I much prefer the term 'computer glitch' as it avoids attributing a human psychological condition to a mere machine. You may of course guess why its producers push for exactly the opposite.

Coming back to Harari's recent essay, he argues that AI has in fact hacked the operating system of human civilisation as it becomes better than humans in mastering language, the foundation of all human culture.

As we are today the guests of one of Europe's best Law Faculties, Harari helpfully reminds us that language is the basis of all law. Until recently, drafting and applying laws required human beings.

However, this is rapidly changing. A recent study by Goldman Sachs claims that 44 percent of all legal tasks can be carried out by AI applications.

In order not to worry aspiring lawyers in the audience, I should hasten to point out AI is likely to enable small and solo firms to take up complex cases which in the past could only have been handled by large firms with huge numbers of staff.

However, it is useful to reflect for a moment what would happen to our societies if our political narratives, laws, literature, visual arts and music would no longer be created - even partly - by humans, but exclusively by AI?

This is not the spectacular sudden end of human civilisation through AI as portrayed in Science Fiction films like 'Terminator' but a gradual , more realistic and yes, frightening scenario.

And of course it is not only language, but also very much images that can be generated by AI, the deepfakes. The ones I mentioned with regard to the Chinese cemeteries a moment ago.

While initial attempts to impersonate politicians, you may remember the Obama and Trump videos, were funny rather than anything else, most recently they have become impressively, or disconcertingly, realistic.

Consequently, it makes sense to think hard about what AI is doing to us as a society as a whole, in addition to what it does to us as individuals. How does the use of AI and automation affect our perception of society

and democracy? Are we prepared for an AI system to be a CEO (already the case in China) or indeed lead a political party?

Before exposing you to further compelling reasons why we should be giving top priority to the regulation of AI through internationally binding legal standards, let me give you a bit of breathing space by dwelling for a moment on the numerous positive aspects AI applications can bring to societies and individuals alike.

Although the vaccines against COVID were not developed by AI, it did assist in the sharing of virtual real time information on research results and the processing of massive data sets.

To stay in the medical field for a moment, dermatologists, to give but one example, can already be greatly assisted by AI to ensure that they have not missed crucial indicators in their diagnostics.

Also in many other areas of healthcare AI is already having a positive impact, which is likely to accelerate.

It will almost certainly help us to better understand the universe, only recently an AI application discovered a star that human astronomers had repeatedly missed.

Transport safety, education and finding ways to tackle the energy crisis are other examples that readily come to mind when considering the positive impact of AI.

And on a more prosaic level, the New York Times reported on 15June this year that AI had assisted Paul McCartney in completing one last Beatles song, to be released, for the fans out here, later this year.

Yet, as the copyright lawyers among you will immediately point out, this last example has already caused ethical and legal controversy.

Unfortunately, ladies and gentlemen, for the purpose of today's argument, we need to leave the benefits of this, many would argue revolutionary, new technology, aside for now and return to its inherent risks.

I have already touched upon some ethical and legal issues other than copyright and referred to some obvious of the shortcomings of recent trendy tools such as Chat GTP, and as well as on the real future gradual risk to our societies. But is that all we should worry about?

The Child Benefits Scandal, *Toeslagenaffaire* in the Netherlands is a perfect example of what can go wrong when governments thoughtlessly use AI systems. The *Toeslagenaffaire* refers to the scandal (or tragedy) in which the Dutch tax authorities used an algorithmic decision-making system to detect fraud. This system turned out to be heavily biased and racist, leading to up to 35,000 people being accused of fraud, which was unjustified in 94 percent of cases. Some 3000 children were taken away from their parents. Following a request by the Dutch Parliament, the Council of Europe's Venice Commission, made up of eminent Constitutional Lawyers, in October 2021 adopted its first-ever opinion on the Netherlands, based on the Parliament's own Report «Unprecedented Injustice ». With regard to the use of Artificial Intelligence by the tax authorities, the Venice Commission observed that discriminatory

practices were systemised through algorithms. It also noted that its findings applied to the Netherlands but might well concern to other countries too.

During the Covid pandemic, disinformation about vaccines spread via AI powered tools, which thus actually contributed to loss of life. As you will all be aware, fake news spreads much faster than the truth – especially if is AI assisted.

Recently, an application similar to Chat GTP recommended suicide to a depressed young man in Belgium.

Tragically, he followed the advice, which led the University of Leuven to send an open letter asking for additional safeguards on powerful interactive systems.

In Italy, as you will probably have followed, the Data Protection Authority challenged Open AI on the protection of minors in Chat GTP and on data protection issues. This led to the temporary voluntary withdrawal of the application from the Italian market.

Several scholars have pointed out that companies spend massively on increasing performance, also known as the AI race, and have reduced their spending on safety. Some have in fact abolished their ethical teams altogether.

On 12 June, the UN Secretary General, Antonio Guterres, expressed his deep concern that AI was becoming a monster of hate and lies because tech firms cared more about engagement than about human rights.

Never in the field of human development has so much power over so many been in the hands of so few.

Another, much overlooked aspect is the sustainability question. The training of AI systems requires huge amounts of energy and thus leaves a big carbon footprint. At Google for instance, it was reported that AI made up for up to 15 percent of the company's total energy consumption, namely a staggering 2.3 terawatt, or 2.3 trillion, watt hours.
It would thus appear very sensible to consider whether an infinitesimal increase in performance or accuracy is worth the equivalent of burning an entire forest.

Another issue of concern is the quality of the data used to train AI models. By 2026 we will have run out of unused fresh data generated by humans.The training of AI models will then have to take place on AI generated data.

However , studies have demonstrated that this « AI incest »leads to degenerative process , hence the term, with increased errors through the overemphasis on popular data , and , ultimately , model collapse.

The military use of AI has been heavily funded for years and for instance fully autonomous killer drones were used for the first time in the recent war between Armenian – Azerbaijan . Needless to say that the Russian aggression agaist Ukraine is giving a further boost to the testing of these lethal tools.

Sadly, arms control negotiations in Geneva on these LAWS (Lethal Autonomous Weapon Systems) have been stalled with not even agreement on the definition of the matter.

Another area of heavy military investment are the brain – machine interfaces. Efforts to enhance the fighting capacity of soldiers are as old as armed conflict itself, and nanotechnology combined with AI is opening up frightening new horizons.

And not only in the military field, Elon Musk's Neuralink company aims to generate brain interface to restore autonomy to those with unmet needs today and unlock human potential tomorrow.

While I have no problem, on the contrary, by assisting those in medical need today, I am decidedly uncomfortable with the aim for tomorrow. How human will it be ?

Well, some of you might immediately retort, what about the open letters we have seen, signed by leading academics and prominent CEO's including Elon Musk and Sam Altman (of Open AI fame) putting the risk of powerful future AI models on a par with pandemics and nuclear holocaust – but not, revealingly, climate change - and proposing a halt to research.

Should we not lose sleep over those?

At the risk of disappointing you, my answer would be no. At least not yet.

Please let me explain why.

In my modest opinion the open letters calling for an, actually only short, pause in the development of powerful AI models pointing at a future cataclysm were disingenuous. First and foremost, they, I am convinced deliberately, the appeal fuelled a hype and drew attention to the products leading companies clearly want us to notice.

Furthermore, a 6 month development pause, even if verifiable, which is a big if, would likely consolidate the comparative advantage of leading companies at the expense of smaller players.

It is hard to see how such a pause could mitigate the risks its proponents are claiming to worry about.

Instead, the companies behind the open letters should address the very real risks that already exist today.

Finally, the letters called for regulation. Altman even mentioned the need to start negotiating an international treaty.

Well, fortunately, responsible European policymakers did not wait for this 2023 appeal but started this work years ago with concrete results about to be delivered in the coming months.

Already in 2018 the Montreal Declaration on the responsible development of AI was officially presented, following a year of multistakeholder dialogue involving government, civil society and companies.

A myriad of other sets of ethical Charters and Guidelines are now in existence as well as Political Declarations prepared by the OECD and

UNESCO, - principles to which Open AI and others could of course adhere to if they really wished to do so.

Meanwhile, European Institutions is currently finalising two major sets of legally binding international AI regulations: the EU AI Act and the Council of Europe Global AI Treaty.

The run-up to the EU AI Act started in 2019 with the EU Ethics Guidelines for Trustworthy AI, which forms the basis of the AI Act proposal.

Also in 2019, the then still 47member State Council of Europe (which expelled the Russian Federation within weeks following its aggression against Ukraine) set up its Ad Hoc Committee on Artificial Intelligence, CAHAI, with a double task: to examine, on the basis of broad multistakeholder consultations, the feasibility and potential elements of a legal Framework for the design, development and application of AI in the field of human rights, rule of law and democracy.

The proposed EU AI Act (covering 27 EU-countries) is a 'product safety regulation' and foresees strong risk-based conditions for AI systems before they can be put on the market.

The European Parliament earlier this month adopted its negotiating position on the Act, which will now be further negotiated in the so-called with the Council of the European Union and the Commission. The aim, which some feel is rather optimistic, is to conclude negotiations by the end of this year.

Let me share with you some of the key elements of the European Parliament's proposals:

Prohibited AI practices

The rules follow a risk-based approach and establish obligations for providers and those deploying AI systems depending on the level of risk the AI can generate. AI systems with an unacceptable level of risk to people's safety would therefore be prohibited, such as those used for social scoring (classifying people based on their social behaviour or personal characteristics). MEPs expanded the list to include bans o
n intrusive and discriminatory uses of AI, such as:

- "Real-time" remote biometric identification systems in publicly accessible spaces;
- "Post" remote biometric identification systems, with the only exception of law enforcement for the prosecution of serious crimes and only after judicial authorization;
- biometric categorisation systems using sensitive characteristics (e.g. gender, race, ethnicity, citizenship status, religion, political orientation);
- predictive policing systems (based on profiling, location or past criminal behaviour);
- emotion recognition systems in law enforcement, border management, the workplace, and educational institutions; and
- untargeted scraping of facial images from the internet or CCTV footage to create facial recognition databases (violating human rights and right to privacy).

High-risk AI

MEPs ensured the classification of high-risk applications will now include AI systems that pose significant harm to people's health, safety, fundamental rights or the environment. AI systems used to influence voters and the outcome of elections and in <u>recommender systems</u> used by social media platforms (with over 45 million users) were added to the high-risk list.

Obligations for general purpose AI

Providers of foundation models - a new and fast-evolving development in the field of AI   would have to assess and mitigate possible risks (to health, safety, fundamental rights, the environment, democracy and rule of law) and register their models in the EU database before their release on the EU market.

Generative AI systems based on such models, like ChatGPT, would have to comply with transparency requirements (disclosing that the content was AI-generated, also helping distinguish so-called deep-fake images from real ones) and ensure safeguards against generating illegal content. Detailed summaries of the copyrighted data used for their training would also have to be made publicly available.

Supporting innovation and protecting citizens' rights

To boost AI innovation and support SMEs, MEPs added exemptions for research activities and AI components provided under open-source licenses. The Act should promote so-called regulatory sandboxes, or real-life environments, established by public authorities to test AI before it is deployed.

Finally, MEPs want to boost citizens' right to file complaints about AI systems and receive explanations of decisions based on high-risk AI systems that significantly impact their fundamental rights. MEPs also reformed the role of the EU AI Office, which would be tasked with monitoring how the AI rulebook is implemented.

It is fair to say that the European Parliament has considerably strengthened the Commission's original proposal by adding stronger human rights protection and by taking a position on very recent developments such as generative AI.

So what is the added value of the other major legislative initiative, the Council of Europe's negotiations on a global Treaty on AI and Human Rights?

Let me take you back in time for a moment.

The Council of Europe was set up in 1949 in response to the horrors of World War 2, in order to protect Europe's common values of human rights, rule of law and democracy.

The Council's approach to meet this huge challenge has been the establishment of binding international treaties, Conventions in the Strasbourg jargon. The best- known is without doubt the European Convention on Human Rights, with its unique judicial supervision mechanism. However, there are in total some two hundred international Conventions. About a third of these, I should add, originated as proposals in the Parliamentary Assembly, thanks to the hard work of Professor Hans Franken and his colleagues.

While these were initially reserved for member States only, the Council of Europe soon realised that they could become global benchmarks.

The Council of Europe also realised very early on that new technological developments would have a massive impact on the values it was set up to protect.

New technologies have been on the agenda of the Council of Europe since the '80s. Consider, for example, the first personal data convention of the Council about 40 years ago, i.e. Convention 108, which can be considered the grandmother of today's GDPR. Or consider the Budapest Convention on Cybercrime from about 20 years ago, which remains the only convention on cybercrime today and now has 68 parties (Nigeria and Brazil were the latest states to join at the end of last year). Capacity building activities are being carried out in some 130 countries.

As it became clear that the development and use of AI by governments pose risks – besides many opportunities – to human rights, democracy and the rule of law, I pushed hard for the Council also to address AI governance.

In September 2019, the Council of Europe established the **"Ad Hoc Committee on Artificial Intelligence" (CAHAI)**, an intergovernmental committee with a two-year mandate from 2019-2021. The CAHAI was mandated to examine the feasibility of a new legal framework for the development, design and application of AI with human rights, democracy and the rule of law as guiding standards. In December 2021, the Committee unanimously decided that there was a need for an additional legal framework on AI governance. In its final report, the CAHAI set out a

checklist of useful elements that should ideally be included in a new binding instrument.

Following this finding, the Council of Europe established the "**Committee on Artificial Intelligence (CAI)**", whose mandate runs from 1 January 2022 to 31 December 2024. The CAI's mandate is extremely ambitious, given that in its two-year existence it aims to elaborate a transversal legally binding instrument in the field of AI in which human rights, democracy and the rule of law prevail. In early 2023, the CAI delivered its first now public draft of an AI Convention, the so-called "Zero Draft".

The AI Convention aims to address the potential risks governments face when deploying AI. It sets out some general principles such as equality, privacy, accountability, transparency, oversight, security, data quality and sustainability that must be ensured at all stages of AI systems (e.g. during design, development and deployment). It is likely to introduces a human rights impact assessment mechanism, regulatory sandboxes and national supervisory authorities.

But, of course, we know that unless everything is agreed, nothing is agreed.

Whereas the proposal for an AI Regulation (the "AI Act") focuses on the European single market and will only apply to 26 member states, the Council of Europe's AI Convention, on the other hand, involves a much larger number of parties and thus aims to have a global impact. Much like Convention 108 and the Budapest Convention.

The negotiating Committee, CAI, includes all EU countries plus an additional 19 member states, including the UK, Switzerland, Norway, and

Ukraine. Moreover, the US, Canada, Mexico, Japan, and the Vatican are Observer States which, together with Israel – at the latter's specific request- also participate in the negotiations. Uniquely, civil society and industry (28 companies and associations which joined the 'Digital Partnership') are at the table as well.

Together with its founder and President, Catelijne Muller, I have the honour to represent ALLAI, an independent organisation dedicated to drive and foster Responsible AI.

As I just mentioned, the negotiations are ongoing and held in camera I cannot, unfortunately, at present share details of the draft Convention.
What I can say, because it is already in the public domain is that there is a debate on th scope of the text, and in in particular whether it should extend to public authorities' use of AI only, or also to the private sector.
You will not be surprised to hear that, loke many others I consider it essential that the minimum safeguards the Convention will provide should be upheld by governments for all uses of AI.

The contrary would be equivalent to insisting that state nuclear facilities need to comply with safety regulations but that private ones do not.
AI is a very powerful technology and should be recognised as such.
I am still surprised to see how little national authorities seem to be aware of this.
Only few states have a comprehensive AI strategy, clear Government policies or a specialised Parliamentary Committee. Or have included AI literacy in the curriculum of their magistrates' training programmes.

So, much needs to be done. And done fast.

To end, I like to return to a quote by Francis Bacon, Britain's first Queen's Counsel in 1597, about money, but which also applies very well to AI systems:

"It is a good servant but a bad master. Either we control it, or it controls us".

Thank you for your attention !

Hans Franken lecture, 30 June 2023

Dear Jan,

I am very happy that you are with us today in this famous classroom, where several Nobelprize-winners – Lorentz and Einstein – lectured in the last century. You gave us a broad view of the work you did in Strasbourg in the Council of Europe.
I had the honour to join the work of the Council being a member of the Parliamentary Assembly from 2010-2015 where I met you in Strasbourg as the head of the Legal Department in the Headquarters of the Council.
You formulated the best possible texts and descriptions with convincing memories of understanding. I remember our meetings always as constructive and efficient with an open eye on the feasibility of the legal documents.
Thank you so much, once again, for being here!

Ladies and gentlemen,
There are many remarks to make about Artificial Intelligence. Stimulated by the entrance of ChatGPT on the market all newspapers publish about it, many with positive and some with negative opinions. But we are sure that in a short time everybody will be confronted with AI.
Just for that reason we can be happy, that our Europe takes the lead to set norms for the use of this phenomenon to stimulate the positive aspects and to eliminate what will be destructive.

As has been told in the lecture of Jan we have two approaches: the Convention of the Council of Europe to stress human rights, the rule of law and democracy. It gives general principles and guidelines for action with the ultimate control of the European Court of Human Rights.

On the other hand the AI Regulation (AI Act) of the European Union with the focus on AI issues from the perspective of market regulation, combined with a set of human rights conditions. It gives a risk-based approach for the systems about what is permitted and which are the conditions for use. In this framework there is a governance structure with supervisors at the national level to look on the activities of the people in the market and
their use of AI systems.

In the light of these two sets of rules there is something missing in our country. We still need an advisory council with independent specialists from the broad

spectrum of the AI applications (human, work, education, care, economics, security etc.) that can give a broad view about the practical and ethical questions where to go, what should be stimulated and what will be dangerous now and in the future - for individual citizens and for the society as a whole. This is much more than supervision only. It will be an essential bridge between the technological world of AI and the political and social world for three activities:
- to assess observations and comments on the AI strategy and to discuss new versions of that strategy,
- to advise the government on data openness
- to assess the impact of AI on industry, government and society.

Such Advisory Councils have been set up already in the US, UK, Canada, Germany, France and Spain.

Very recently a small group of specialists came together to realize this goal for our country. I hope they will have success in their dialogue with the cabinet that has started this week.

Last but not least I would like to thank the eLaw Board, especially Bart Custers and Regina Noort for organizing today's happening, which is connected with my name. And I am happy to know that also many students (most of them online) are joining us.

Thanks to all of you for being here, be it in person or online. Together we shared the development of a new topic that will become part of our life!

Hans Franken